Research Paper | Computer Science | E-ISSN No : 2454-9916 | Volume : 10 | Issue : 4 | April 2024

SPEECH-TO-TEXT TRANSCRIPTION FOR CORPORATION GRIEVANCE CALLS

Sampath MK[1], Sujeet P[2], Sanjay P[3], Sanjay S[4], Soundarrajan A[5]

[1] Assistant Professor, Computer Science and Engineering, Knowledge Institute of Technology, Salem, India.
[2, 3, 4, 5] UG Students, Department of Computer Science and Engineering, Knowledge Institute of Technology, Salem, India.

ABSTRACT

To evaluate and enhance an existing open-source speech-to-text transcription tool for accurately converting feedback calls about citizen grievances into English text. The focus lies on achieving measurable improvements in transcription accuracy across calls made in Tamil or English. Rather than creating a new system, the project concentrates on refining an established open-source solution. The endeavor spans domains including Natural Language Processing (NLP), speech recognition algorithms, and machine learning. Through this effort, the objective is to enhance the tool's efficacy in processing multilingual data and better serve the needs of diverse linguistic communities in citizen grievances.

KEYWORDS: Natural Language Processing, Speech Recognition, feedback calls, Citizen grievances, linguistic communities.

1. INTRODUCTION

If you're looking to develop a system for handling complaints from citizens who speak different languages, it might be worth considering the use of Natural Language Processing and Speech Recognition technologies. These tools can help ensure accurate transcription and interpretation of feedback calls, even when the speaker has a non-standard dialect or accent. By using these technologies, you can improve communication and resolution of issues, which will ultimately lead to better community engagement and satisfaction. Keywords: Natural Language Processing, Speech Recognition, feedback calls, citizen complaints, linguistic diversity. This innovative project is designed to revolutionize the way complaints are lodged with government agencies and corporations in Tamil Nadu, especially catering to those who are not proficient in Tamil. By introducing a user-friendly platform that supports multiple languages, it aims to make sure every resident's voice is heard, regardless of their linguistic background. Utilizing cutting-edge speech recognition and translation technologies, the system simplifies the complaint submission process: users can easily record their grievances in their preferred language. These recordings are then accurately transcribed and translated into Tamil, ensuring that every complaint is clearly understood and effectively processed.

The initiative's core mission is to break down language barriers, fostering a more inclusive and transparent interaction between the populace and authorities. It empowers individuals from various linguistic groups, enabling them to communicate their concerns to governmental bodies and corporations without the fear of miscommunication. By integrating advanced speech-to-text and translation functionalities, the platform guarantees the precise conveyance of every complaint, enhancing the overall efficiency and responsiveness of governance. This approach not only improves the user experience but also plays a crucial role in promoting a governance framework that is more accountable and responsive to the needs of its diverse population.

2. LITERATURE REVIEW

In the evolution of automatic speech recognition (ASR) research over the past decade, the integration of deep learning techniques has led to surprisingly significant changes, with more than 50% reductions in single-word errors compared to unused implementations. deep learning. This significant advance is being driven by the development and adoption of an all-neural, end-to-end (E2E) ASR model that represents a leap forward towards multi-level integration and joint venture. Unlike previous models, the E2E model leverages the holistic use of general machine learning models, making it possible to learn directly from data without relying on complex, specific knowledge domains that were previously important for ASR systems. The development of the E2E model, which includes all neural architectures, demonstrates a general trend toward simplicity and efficiency in ASR research, making this model state-of-the-art in the field.

This review aims to provide a detailed overview of E2E ASR models, showing their evolution, key features, and how they compare to classical Hidden Markov models, the model (HMM)-based framework that once dominated ASR technology. This project provides an in-depth study of the E2E ASR process, covering a wide range of topics from modeling, training, and decision-making strategies to sharing external language models. It not only shows performance metrics and actual deployment conditions but also takes into account the future directions of the technology. The discussion highlights the evolution of deep learning in ASR, showing how these advances can improve processes, improve performance, and expand the application of knowledge management.[1]

In automatic speech recognition (ASR), neural sensors play an important role by improving the accuracy and efficiency of converting speech into text. Building on this progress, this paper describes the use of neuro transducers in end-to-end speech stream (ST), a state-of-the-art method designed to directly convert music into articles in many languages. This new approach uses a Transformer (TT)--based ST model, which reduces the expected latency compared to the cascading ST approach based on ASR and post-text translation technology (MT). The TT-based model has an advantage: it not only improves speech quality but also reduces the risk of error propagation from the ASR stage to the MT stage, a common problem in cascaded systems.

To improve the performance of the model, this article introduces integration into the coordination of the TT framework, a new tool designed to improve the model's ability to solve complex language patterns and nuances. Additionally, the TT-based ST model extends its functionality to multiple ST languages, making text available in multiple languages simultaneously, a feature that is very useful in today's world. Analysis of a 50,000-hour pseudo-labeled dataset shows that the TT-based ST model not only reduces inference time but also exceeds the performance of traditional flowless cascade ST models, especially in Anglo-German hands. This breakthrough has led to tremendous progress in making multilingual communication easier and more efficient, with the ability to transform interactions across multiple languages and relationships around the world.[7]

While more than 85 percent of languages in Canada are considered negative, preserving and revising these languages is critical to preserving Aboriginal knowledge and cultural heritage. To support these efforts, there is an urgent need to develop specialized computing tools that will help language communities and linguists work on the knowledge and support of words. One of the biggest problems encountered in this process is the prevalence of different languages in spoken language.

To solve this problem, we propose a comprehensive pipeline designed to effectively process multilingual data. Our approach combines customized training and business speech-to-text services to provide a variety of solutions suitable for many language environments. Our approach differs from the user format, specifically the well-known ELAN validation practice created by Brugman and Russel in 2004. This integration not only facilitates ease of use but also improves access to community programs and academic research.

Using this pipeline, the language community and language teachers can access powerful tools that simplify the knowledge process and promote the language. The combination of educational models and business services leads to the accuracy and description of speech data in multiple languages, making it easier to analyze language and information. Additionally, the relationship with ELAN enhances the results of our approach, making it possible to integrate it into existing projects and methods used by language researchers and community members.

Overall, our proposed pipeline is an important step in promoting language support in Canada and beyond. By providing easy-to-use and effective tools, we aim to improve collaboration between language communities and language teachers to preserve and restore Aboriginal languages, thereby promoting cultural heritage and linguistic diversity.[10]

This article introduces the Multilingual Libri Speech (MLS) dataset, an important resource in speech science. Based on the audiobooks available at LibriVox, MLS provides a broad and diverse framework suitable for the study of many aspects of speech in different languages. The collection consists of recordings in eight languages; of these, approximately 44,500 hours are in English, with the remaining languages accounting for approximately 6,000 hours.

MLS is unique not only for its scale but also for its language models (LM) and automatic base speech (ASR) models for all languages included in the documents. This comprehensive course provides researchers with the tools necessary to teach multilingual knowledge and text-to-speech (TTS) production in multiple languages.

By providing access to such a vast collection of transformative data, MLS has the potential to support innovation and open new avenues for ASR and TTS research. Researchers can use MLS to train and evaluate speech recognition in different languages; so they can solve problems related to different languages, dialects, and dialects. Additionally, the availability of LM and ASR test patterns can serve as important benchmarks for evaluating the effectiveness of new techniques and methods.

Overall, MLS represents an important resource for the speech community by supporting the development of multilingual ASR and TTS technologies. Its accessibility and comprehensibility make it the basis for future research aimed at improving speech recognition and communication in multiple languages.[3]

This article proposes ways to overcome the limitations of multilingual speech-to-speech (TTS) systems, especially on low-language data. While neural TTS has made significant progress in creating human-like language communication, current multilingual TTS generally produces rich language and relies on text links and Good audio equipment. The plan proposes a revolutionary change by enabling zero-throw multilingual TTS using only information in that language.

This data-only approach democratizes TTS development, especially for weak text-only data. This breakthrough makes TTS technology available for thousands of languages that were previously excluded due to a lack of data. The framework is inspired by the evolution of different languages in multilingual models, starting with masked language models before training using only multilingual data. The model is careful to use connections while maintaining a stable grammatical layer.

This new approach makes it possible to understand words

that are not included in the linked data but only in the text. Evaluation results demonstrate the effectiveness of the method, demonstrating a good understanding of zero-drop TTS performance. Remarkably, the character error rate for invisible words is still very low, below 12%.

This article expands on the use of TTS technology for multilingual conversation between internationals, going beyond reliance on text and audio files. The ability to create accurate and intelligent contact words for rare languages is beneficial for communication, education, and cultural preservation. Additionally, this approach demonstrates the potential for using advanced language models to solve real-world problems and encourage participation in technological development.[5]

This survey provides a comprehensive evaluation of automatic speech recognition (ASR), an area that has received great attention due to its many applications. This article begins with an in-depth examination of the core concepts and theoretical foundations of ASR, aiming to demystify the principles that govern the discipline. He then began analyzing various end-to-end speech recognition systems to understand their strengths and weaknesses. With a special focus on Connectionist Temporal Classification (CTC), RNN-Transformers, and Transformer-based models, their functionality and efficiency are explained.

Furthermore, this article highlights the unique challenges faced in: ASR for Indian languages offers a unique landscape of diverse languages and resources. In the discussions, the important challenges and great opportunities that this diversity offers, especially in the field of science, were emphasized. The main issues highlighted include the huge gap between English and Indian languages and the lack of teaching materials that hinder the development and implementation of effective ASR systems.

The main objective of this survey is to provide readers with an understanding of ASR, its development, and future potential, with special emphasis on the subject of the Hindi language. Providing an in-depth understanding of the progress of ASR research and describing the current challenges and opportunities in meeting India's diverse needs, this article focuses on researchers and professionals in the field of decision-making and new solutions in speech recognition.[9]

This article describes a new method for developing a code-switching (CS) multilingual automatic speech recognition (ASR) system that can record messages containing many different languages. First, it overcame the challenge of achieving data transfer by introducing a new way to generate CS ASR data using only a single language data. This approach expands the availability of training data for CS ASR models, increasing their efficiency and effectiveness in both languages.

Second, this article introduces a pioneering technique called tandem tokenizer; This technique is designed to support the CS ASR model to generate a message for each output token when using tokenizers in a language that already exists. This cascading tokenizer not only facilitates the integration of language recognition and ASR processes but also optimizes the use of available resources and increases efficiency without compromising reality.

The effectiveness of this method has been proven by testing two languages, English-Hindi and English-Spanish, using Miami Bangor CS corpus analysis. The results reveal a new level of cutting-edge technology achieved by the proposed methods and demonstrate the accuracy and precision of these methods in speech processing.

Furthermore, the cascade tokenizer model achieves more than 98% data accuracy from FLEURS, demonstrating superior information in language recognition. This demonstrates the efficiency and power of the proposed method not only in writing code to manipulate speech but also in recognizing correct speech and thus extending its benefits beyond the activities of ASR.

Overall, this research improves the scope of ASR projects. Jen's multilingual ASR system provides practical solutions for building datasets and tokenization technologies, improving the functionality and performance of different languages, and ultimately supporting the advancement of speaking multiple languages.[4]

3. EXISTING SOLUTION

The system must now use open-source speech-to-text tools or libraries as its foundation. Hindi should have a simple support system for multiple languages, including English, and should be bilingual, reflecting the combination of Hindi and English commonly used in India. More importantly, the system can detect beginner-level accuracy, and the performance test is designed for typing words in different languages. These measures are intended as a starting point for evaluation and optimization and as the basis for further development. The current system is intended to provide a functional role for speechwriting tasks, but there is room for refinement and refinement to suit the nuances and complexity of speech of many words. The system has laid the foundation for continuous improvement to increase accuracy, efficiency, and integrity across multiple languages by creating an open platform and leveraging existing language support.

4. PROPOSED SOLUTION

The solution addresses the current data gap in Tamil Nadu through various methods, using technology to increase accessibility and operate efficiently and accurately. First, a comprehensive assessment and evaluation process will be conducted to review the effectiveness of existing tools and identify areas ripe for improvement. This will be followed by careful data collection focused on collecting large amounts of audio data containing the different languages and voices of residents. > Additionally, speaker classification algorithms will be used to identify speakers and sections in the system, while the use of advanced language recognition techniques will help support more languages.

Seamlessly integrating components, these solutions are built to

simplify the process of submitting complaints by overcoming language barriers that hinder effective communication between citizens, clans, and police. This integration not only improves the accuracy of data collection but also speeds up the complaints process, ultimately ensuring the right to respond to insiders' concerns.

Solutions required to transform data through the integration of technology and data-driven approaches are not satisfactory in Tamil Nadu. By promoting accessibility, efficiency, and accuracy, it paves the way for a more comprehensive management system that can adapt to the different needs of the people it works with.

5.METHODOLOGY

Data Collection:

The main focus of our efforts is to be vigilant in collecting the majority of the company's unwanted calls. This file will cover many scenarios, including different speakers, different sounds, and different background noise levels. This comprehensive program ensures that language skills are trained in different situations around the world, increasing accuracy and improving general abilities. By integrating various situations and features into our data, we strengthen the foundation of speech recognition and make it more useful in solving complaints handling challenges in the business environment.

Preprocessing:

In the preprocessing phase, the aim is to improve the quality of the collected data to aid clear feature extraction. This includes removing distortion, background noise, and distortion to ensure clarity and consistency in the recording. Technologies such as noise reduction and filtering are used to improve overall sound quality. A normalization method was also used to make sound levels the same across all recordings, thus reducing differences that could affect subsequent analyses. These preliminary steps, by carefully cleaning and enhancing audio files, provide a solid foundation for extracting key elements that are important for creating strong words and acoustic patterns.

Acoustic Model Training:

Training acoustic models involve the use of deep learning techniques such as deep neural networks (DNN) or convolutional neural networks (CNN). This model works by examining the graph of acoustic features extracted from speech symbols and their corresponding words. Thanks to extensive training, the model can make the most of the relationship between speech and content by recognizing phonemes or subwords. Using the power of deep learning, acoustic models identify subtle patterns in speech data, allowing them to accurately record speech and facilitate various speech processing processes precisely and efficiently.

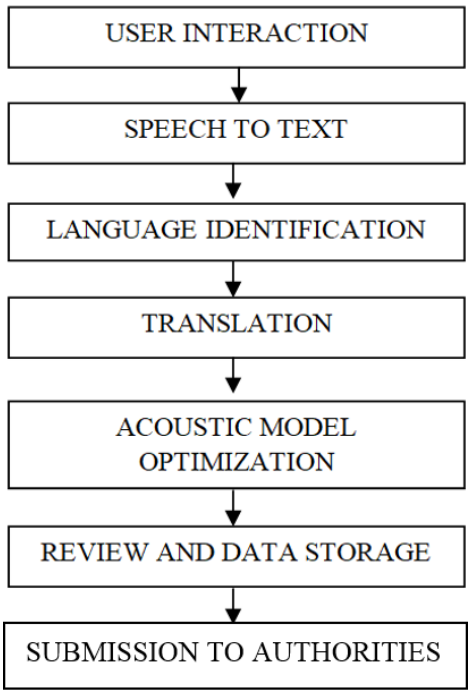
Language Model Integration:

The combination of good language patterns improves the recognition of truth by using the possibility of a temporary word in the context of corporate complaints. Use a variety of techniques such as the N-gram model, Relational Neural Networks (RNN), and Transformers to beautifully capture language patterns and expressions in conversation. By using this format, the quality of the written text is improved, ensuring that complaints are recorded and fully understood. N-gram models help analyze consecutive words while RNNs and Transformers are good at capturing remote control and different contexts on speech devices. Integration of these language models not only increases the accuracy of typing but also increases the overall effectiveness and efficiency of complaint handling in a corporate environment by encouraging communication and problem-solving.

Continuous improvement:

Continuous improvement of the system requires a dynamic process of collecting user input and monitoring performance. By asking for input from users and taking a closer look at how the system works in a real-world environment, areas for improvement can be identified. This revision includes adding additional information and changing the structure according to changing user needs and speech patterns. Thanks to this continuous optimization process, the system maintains its efficiency and adaptability over time. Continuous improvement by responding to user feedback and incorporating new insights allows the system to adapt as needs change and technology advances, ensuring good results and customer satisfaction in the long run.

6. PROCESS DIAGRAM



1. User Interaction:

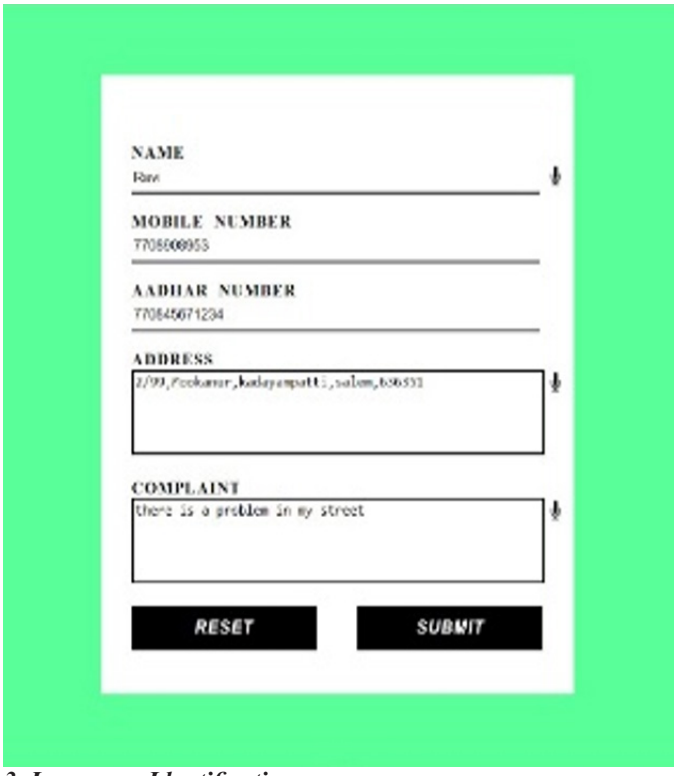
Effective customer interaction is crucial when making phone calls and requires good listening, thinking, and clear communication skills. By listening carefully to callers' concerns, agents can demonstrate empathy and build rapport, creating a sense of support. Clear and concise communication is crucial to understanding current issues and providing timely assistance. Additionally, understanding callers' emotions and acknowledging their concerns helps create positive interactions that increase customer satisfaction and trust. By prioritizing

these elements in user interactions, businesses can resolve complaints, strengthen customer relationships, and manage their chains. Renowned for responsiveness and -compassionate service.

The site is in Tamil, a language spoken in southern India. The login form has fields for username and password. There is also a button labeled “Login.” There is a language selection menu under the login form. Users can choose between English, Tamil, and Hindi.

2. Speech to Text:

Speech-to-text technology transforms communication by converting spoken words into text, making it easier to transfer between unwanted calls. This new tool supports the information process and analyzes and resolves caller concerns. Speech-to-text technology improves accessibility and understanding by accurately converting speech to text, ensuring complaints are accurately recorded and understood. This not only improves the problem-solving process but also allows organizations to gather valuable insights from continuous improvement calls. Overall, speech-to-text technology enables organizations to better serve their stakeholders by providing clear, concise, and effective communication.



3. Language Identification:

Language recognition plays an important role in managing multilingual calls. Machines with language recognition features facilitate translation and effective communication by quickly identifying the words spoken by callers. This is important to ensure that the complaint is well understood and resolved quickly, regardless of the caller’s background. By implementing this process, organizations can improve their complaint resolution processes, making them more efficient and responsive. Language recognition enables systems to automatically translate into different languages of callers, thus promoting collaboration and enabling effective human communication and resolution of outstanding problems.

4.Translation:

Translation services play an important role in overcoming complaints during emergency calls and ensuring effective communication between callers and speaking representatives. Accuracy of interpretation is essential to ensure that complaints are understood and resolved in a timely and effective manner. Translation services communicate differently, allowing people to share concerns and find solutions regardless of language. This creates a more integrated and efficient environment for resolving complaints, improving communication, and increasing overall satisfaction for callers and agents.

5. Acoustic Model Optimization:

Acoustic model optimization is an important process in improving speech recognition, especially when dealing with unwanted calls. By adjusting these patterns, the system will better understand the variety of voices, speech patterns, and background noises often seen during calls. This optimization ensures that complaints are considered and understood, resulting in more efficient and effective writing. Additionally, acoustic structure optimization plays an important role in

timely and effective responses to complaints by reducing the impact of changes in speech and environmental characteristics, ultimately improving the overall outcome of the complaint resolution process.

6. Review and Data Storage:

Examining complaint calls and storing security information plays an important role in the organization's work. Provides detailed analysis of call details to identify emerging issues and trends. Additionally, the process ensures compliance with regulatory processes by maintaining confidentiality and accountability standards. Legal information management processes protect sensitive information and preserve trust between the organization and its stakeholders. By carefully reviewing data and implementing good retention practices, organizations can obtain valuable feedback from complaints, resolve issues quickly, and support a culture of transparency and accountability.

7. Submission to Authorities:

The right to report complaints to authorities, whether the regulator or the compliance group, is the basis of compliance management and transparency of company procedure. This important action facilitates the investigation and resolution of complaints at a higher level. By providing relevant information to relevant organizations, organizations can monitor the quality of environmental management while creating an environment of accountability and increasing stakeholder trust. Publication systems not only ensure compliance with regulatory requirements but also improve management standards by promoting a culture of ethics and responsibility. Respect and maintain the reputation of the organization.

7. CONCLUSION

In summary, the development of the complaint form is an important step towards effective, transparent, and efficient governance in Tamil Nadu. The platform combines technologies such as speech recognition, translation, and data security to enable residents to voice their complaints in the language of their choice rather than saying they speak Tamil. A combination of speech-to-text, language recognition, and translation algorithms to ensure accuracy and confidence in typing. Additionally, improvements such as acoustic model optimization and speaker assignment improve the output quality.

The platform's intuitive interface and comprehensive analysis allow users to participate in the complaints process, promoting civil society, cooperation, and accountability. Ultimately, the whole point of the program is to eliminate the language barrier and promote governance by allowing residents to easily voice their concerns and concerns. By bridging the gap between residents and authorities, the platform will not only enable better communication but also promote a better understanding of community participation and responsibility, thereby contributing to the overall health and well-being of Tamil Nadu. Management responsibility.

REFERENCES

1. Handalage, Upulie, et al. (2021) "Computer Vision Enabled Drowning Detection System", 3rd International Conference on Advancements in Computing (ICAC). IEEE, pp. 240-245.
2. A. I. N. Alshbatat, Abdel Ilah N and Alhameli (2020) "Automated Vision-based Surveillance System to Detect Drowning Incidents in Swimming Pools", Advances in Science and Engineering Technology International Conferences (ASET), pp.
3. Alotaibi, A. (2020) "Automated and intelligent system for monitoring swimming pool safety based on the IoT and transfer learning", Electronics, Volume 9, pp. 2082.
4. Eng, How-Lung and Toh. (2003) "An automatic drowning detection surveillance system for challenging outdoor pool environments", Computer Vision, IEEE International Conference on. Vol. 2.
5. Hayat, Muhammad Aftab and Yang. (2019) "The Swimmers Motion Detection Using Improved VIBE Algorithm", International Conference on Robotics and Automation in Industry (ICRAI), pp. 1-6.
6. Lu, Wenmiao, and Yap-Peng Tan. (2004)"A vision-based approach to early detection of drowning incidents in swimming pools", IEEE Transactions on Circuits and Systems for Video Technology, pp 159-178.
7. Roy, Ajil, and K. Srinivasan. (2018) "A novel drowning detection method for the safety of swimmers", 20th National Power Systems Conference (NPSC), pp.1-6.

S.no	Input	Proposed Performance	Google Performace	Performance (Proposed Vs Google) Yes/No
1	Our Street Light Are Not Functioning Properly	எங்கள் தெருவிளக்கு சரியாக இயங்கவில்லை	எங்கள் தெருவிளக்கு சரியாக இயங்கவில்லை	Yes
2	There Is No Water Facility In Our Street	எங்கள் தெருவில் தண்ணீர் வசதி இலலை.	எங்கள் தெருவில் தண்ணீர் வசதி இல்லை.	No
3	There Is A Electricity Problem In My Area	என் பகுதியில் மின்சாரப் பிரச்சினை உள்ளது	என் பகுதியில் மின்சார பிரச்சினை உள்ளது	Yes
4	Our Street Drainage System Is Not Working Properly	எங்கள் தெருவின் கழிவுநீர் அமைப்பு சரியாக சயெல்படவில்லை	எங்கள் தெரு வடிகால் அமைப்பு சரியாக இயங்கவில்லை	Yes